
PDS-D: Advanced Data Distribution for a Heterogeneous, Geographically Distributed Archive

**JPL IT Symposium 2002
Steve Hughes and Joel Wilf**

November 4, 2002

PDS-D



- **Challenge:** To integrate a collection of resources and provide seamless access to distributed data repositories
- **Solution:** A system that is fully functional, operationally reliable and extensible, and is capable of:
 - Adapting to mission(s) having more complex payloads and significantly larger data volumes.
 - Adapting to not yet compliant PDS data sets.
 - Providing unified web based search-and-retrieval user interface to novice and sophisticated users.
 - Using Internet as the primary method of data distribution.
 - Supporting real time (on demand) distribution of data to users.
 - Provide ability for users to subscribe for notification of released data.

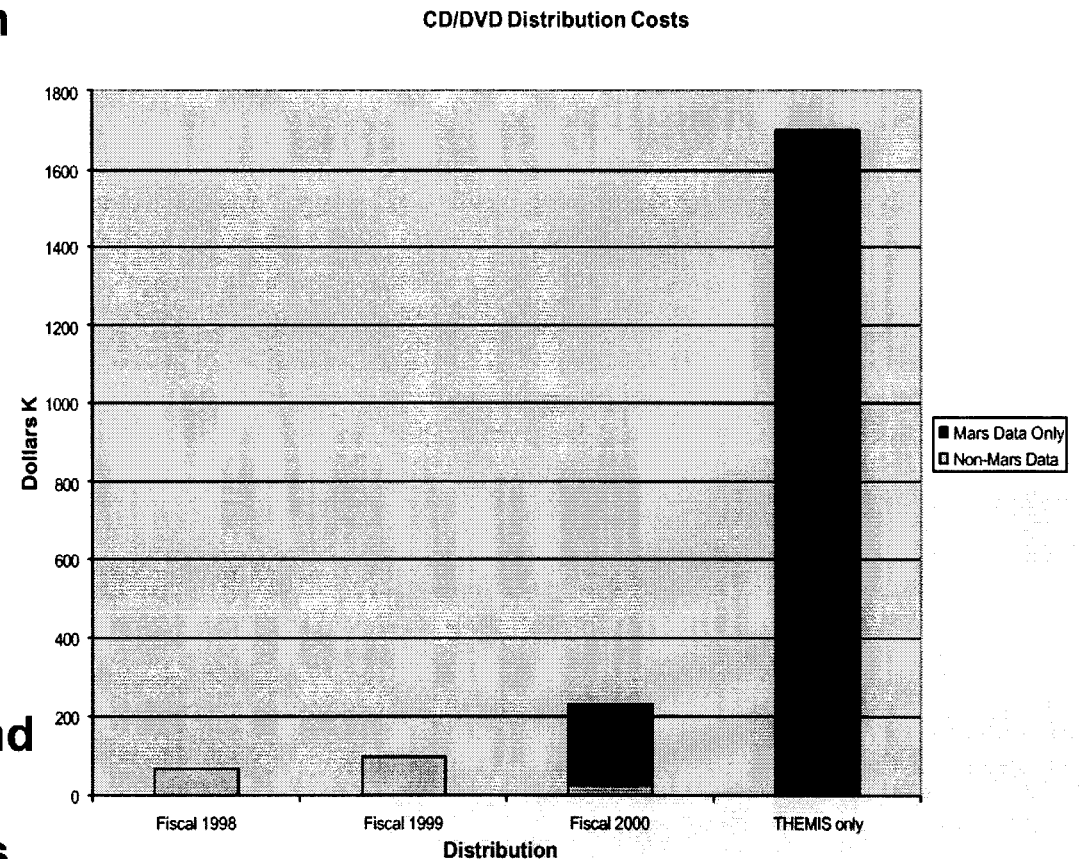
Why a new development?



- Increased data volumes too much for PDS CD distribution system

- MGS CD costs exceeded \$500K for 600GB
- THEMIS alone will produce over 4TB of data
- CD/DVD distribution too costly and cumbersome

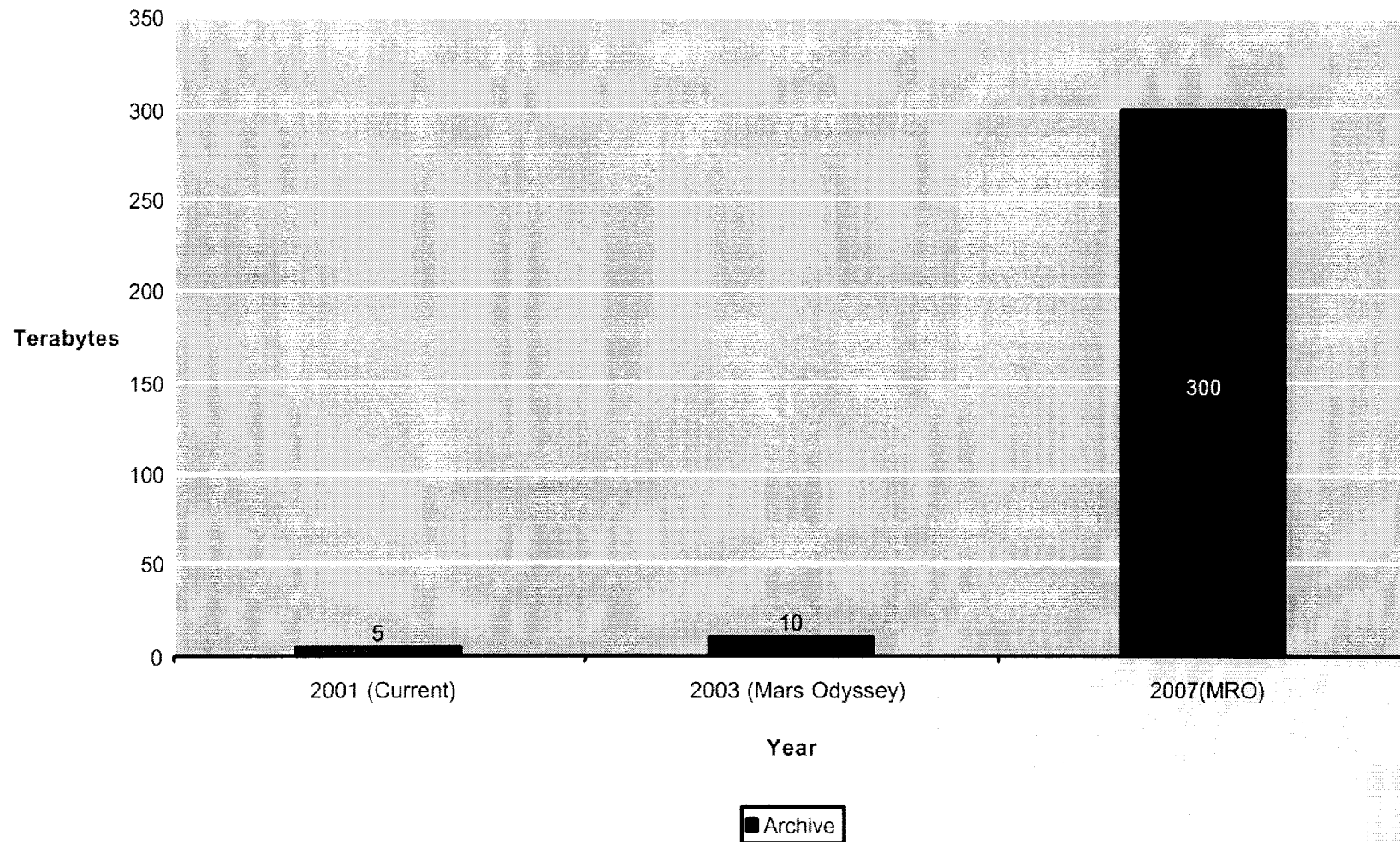
- Data will be physically distributed among various sites and sources
- Data will be more complex and heterogeneous
- Scientists want online access to all available data



Mars Exploration Program: Very Large Data Volumes



Growth in Size of PDS Archive

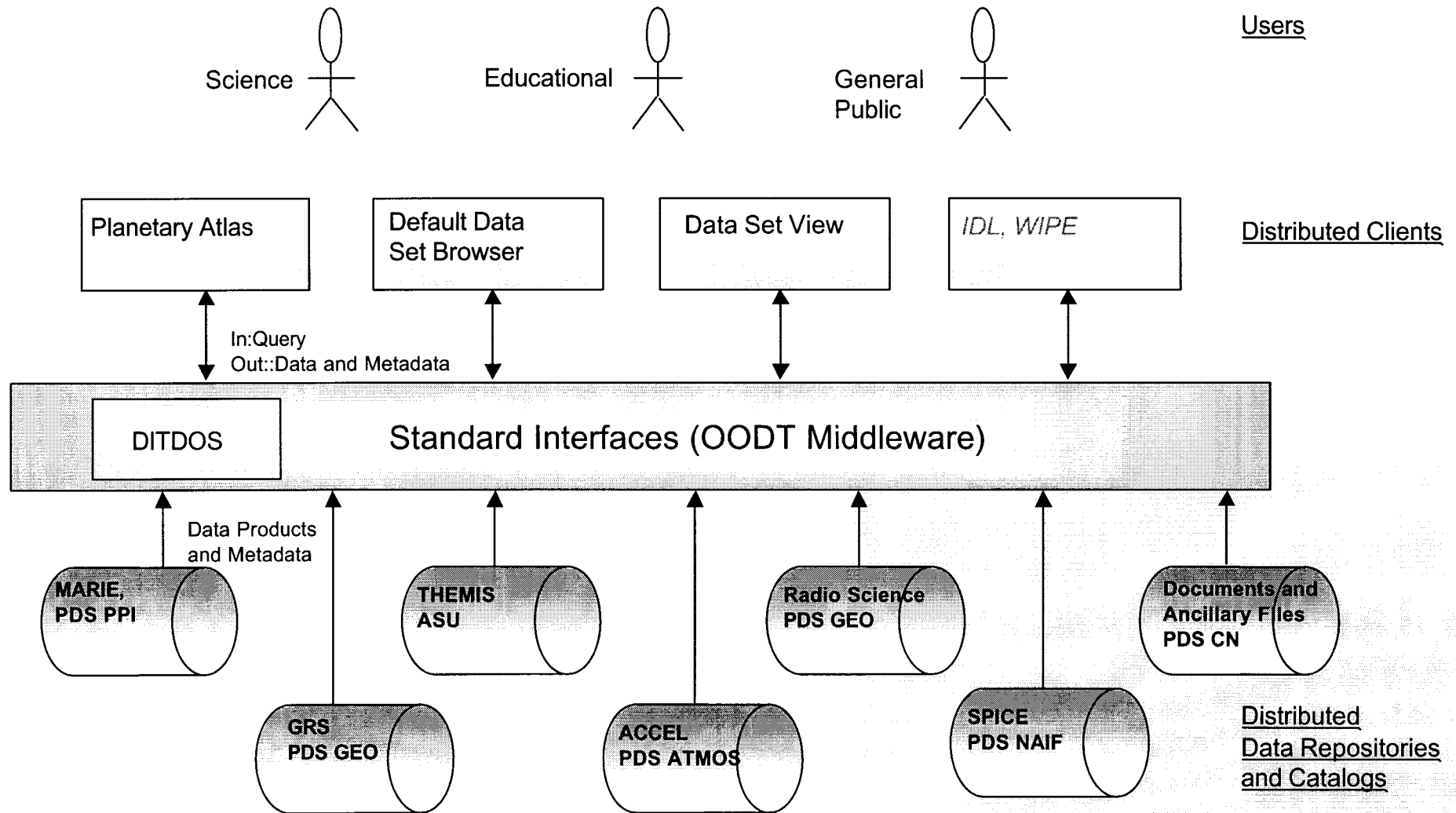


PDS-D Proof-of-Concept D01



- **Provides on-line search and retrieval as the primary method of data distribution**
 - Provides correlative search across instruments
 - Data Set Search
 - Data Product Search
 - Provides simple product retrieval from distributed data repositories
- **Provides unified search and retrieval system interfaces**
- **Makes data publicly available as soon as possible**
- ***Allows physical media copies by request***
- **Focuses on the distribution of Odyssey data products**

PDS-D D01 Architecture - Mars Odyssey



D01 Distribution Requirements Odyssey Data Sets



THEMIS raw infrared data (EDR)
THEMIS raw visible data (EDR)
THEMIS calibrated infrared data (RDR)
THEMIS calibrated visible data (RDR)

ODY-M-THM-2-IREDR-V1.0
ODY-M-THM-2-VEEDR-V1.0
ODY-M-THM-3-IRRDR-V1.0
ODY-M-THM-3-VISRDR-V1.0

GRS raw data (EDR)
GRS binned signal strength (RDR)

ODY-M-GRS-2-EDR-V1.0
ODY-M-GRS-4-BIN-V1.0

MARIE raw energetic particle data (EDR)
MARIE reduced energetic particle data (RDR)

ODY-M-MAR-2-REDR-RAW-DATA-V1.0
ODY-M-MAR-3-RDR-CALIBRATED-DATA-V1.0

Radio Science data

ODY-M-RSS-1-RAW-V1.0

SPICE data

ODY-M-SPICE-6-V1.0

Accelerometer EDR data
Accelerometer profile data
Accelerometer altitude data

ODY-M-ACCEL-2-EDR-V1.1
ODY-M-ACCEL-5-PROFILE-V1.2
ODY-M-ACCEL-5-ALTITUDE-V1.1

Requirements



- **Support a single web-based-user-interface entry point that provides online access to all D01 resources**
 - User interfaces
 - Catalogs and indexes
 - Data repositories (e.g. disk, volumes on-line)
 - Value added applications (e.g. on-the-fly processing)
- **Support the search of**
 - Data product information and location
 - Data products and ancillary data
 - Correlative searches across data sets
 - Data set, documents, and associated information
- **Support the retrieval of**
 - A pointer to a data product file or resource (URL)
 - The data product (e.g., a binary file)
 - The data product label
 - A package containing the data product and associated files
- **Provide unified system interfaces for search and retrieval across distributed heterogeneous catalogs and data repositories**

Requirements (cont)



- **Have minimal impact on existing PDS resources**
 - Based on the existing PDS archive requirements
 - Imposes no new archive requirements
 - Adds a light set of distribution
- **Middleware user support**
 - <http://oodt.jpl.nasa.gov/>
- **Customer base**

PDS-D D01 Implementation

What we did



- **Implemented a multi-tiered information architecture**
 - Application Clients (Browsers/Interfaces)
 - Middleware (OODT)
 - Data and Metadata Servers (product server, profile server)
 - Data Repositories and Catalogs
- **Simplified and *standardized system interfaces* through middleware**
- **Used existing PDS subsystems but *hid heterogeneity***
 - User Interfaces (Atlas, DITDOS)
 - Data repositories (disk farms)
 - Catalog databases (Sybase, Gatesware,...)
 - Remained geographically distributed and locally managed
- **Separated Data Architecture from Technology Architecture**
 - Used archive *metadata* to its full potential
 - Evolved technology architecture

Data Product Retrieval



Product

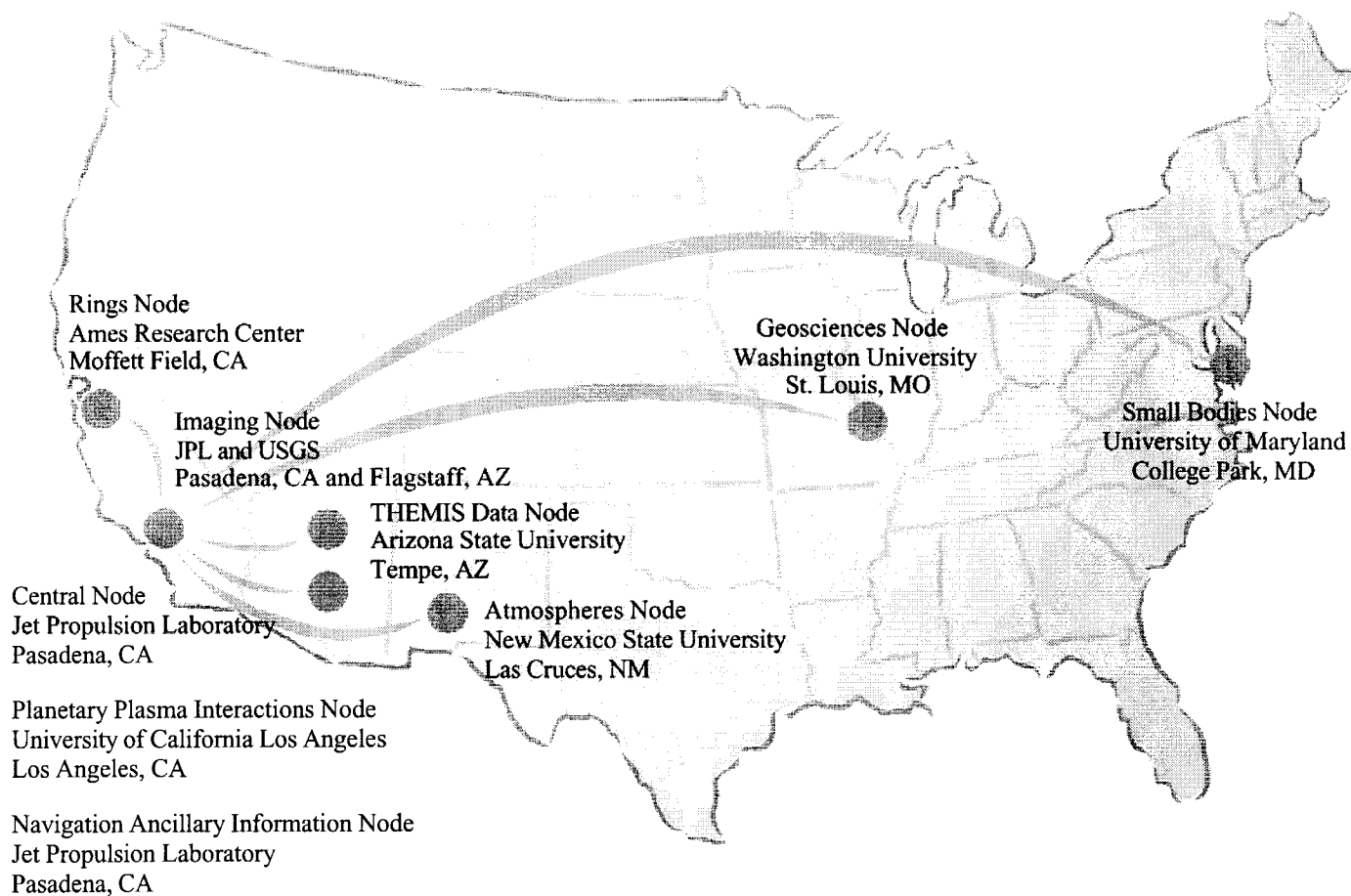
THEMIS
GRS
MARIE
SPICE
ACCEL
RS
Backup

Product Servers

PDS.ASU.Product
PDS.Geo.Product
DITDOS
PDS.NAIF.Product
PDS.Atmos.Product
PDS.RS.Product
PDS.CN.Product



Planetary DataSy stem Distributed Planetary ScienceArc hive



The Product Server in a Nutshell



If a product server has read privileges to a file, it can return that file.

If a product server has read privileges to a directory, it can return all files in the directory, packaged as a zip file.

If a product server has read privileges to a PDS labeled product, it can return all files referenced within the label of the product, packaged as a zip file.

The PDS/OODT product server is capable of serving the vast majority of all products in the PDS archive. (I.e. The product server is not constrained by any target body, mission, data set, or the data repository layout.)

The currently released product server is installed at six nodes. (I.e. All product server capabilities are at all nodes.)

The Product Server in a Nutshell (cont)

Query Handlers



Return_Types

PDS_LABEL – return PDS label

PDS_ZIP – return PDS labeled file and all associated files in a ZIP package

PDS_ZIPN – same as PDS_ZIP except for 1-n PDS labeled files

RAW (mime_type) – return specified file

DIRLIST – return list of all files in a directory

PDS_ZIPN_TES – returns TES product in a ZIP package

PDS_JPEG – convert PDS image to jpeg

Under consideration

PDS_CSV – convert PDS binary TABLE to common separated value ASCII file

PDS_PDS – Normalize data representation of a PDS product

PDS_FITS – Convert PDS product to FITS

...

<http://buttons.jpl.nasa.gov:9002/index.html>

Data Product Search



Product

Component

THEMIS

Planetary Atlas

GRS

Planetary Atlas

MARIE

DITDOS

SPICE

Default Data Set Browser

ACCEL

Default Data Set Browser

RS

Default Data Set Browser

Correlative Search

Planetary Atlas

Data Set/Resource

PDS.CN.Profile (profile server)

The Profile Server in a Nutshell



- **Given search parameters that exist in the PDS model (e.g. TARGET_NAME = IO, FILTER_NAME=RED, RESOURCE_CLASS=DATA.DATA_SET) a profile server is able to return the description and location of all entities that match those parameters.** (I.e. The profile server is constrained only by the existing model.)
- **The currently released profile server is installed at the central node with access to the data set catalog**
 - Searches and retrieves data set information and location of primary data set browser
 - Searches and retrieves resource information and its location (e.g. web site, volumes online)

Data Repositories



Product

Location

THEMIS

ASU

GRS

Geosciences

MARIE

PPI

SPICE

NAIF/CN

ACCEL

Atmos/CN

RS

Geosciences/CN

High Level Catalog

CN

Archived Data Sets

Pending Data Sets

Releases

Resources

System Administration



- **Catalog Management**
 - Archive Data Sets
 - Pending Data Sets
 - System Resources
- **Data Set Release Object**
 - Describes a data set or portion of a data set just released
- **Subscription/Notification Service**
 - User interface for subscribing to release notifications
 - Notification email sent to subscribers
- **Server Manager (Distributed System Management)**

Demonstration

How PDS-D D01 works

Data Set View



Data Set Quick Search

Select ONE parameter from below to perform your query.

RESET

Missions:

2001 Mars Odyssey
Mars Global Surveyor
Mars Observer
Mars Pathfinder

Target Name:

All

Target Type:

All

Instrument Name:

Accelerometer (ACCEL)
Camera 1 (CAM1)
Camera 2 (CAM2)
Gamma Ray Spectrometer/High Energy Neutron Detector (GRS)
JPL Infrared Spectroscopy Data
Imager for Mars Pathfinder (IMP)

Instrument Type:

All

[Advanced Search](#) [Power Search](#)

GO!

<http://starbrite.jpl.nasa.gov/pds>

Data Set View – Results





PLANETARY DATA SYSTEM



[New Search](#)
[About](#)
[Technical Help](#)
[Feedback](#)
[Home](#)


Search Results (5 data sets found)

Data Set	Instrument Host	Information About the Data Set	Data Products & Related Files	Other Resources
1. 2001 Mars Odyssey Gamma Ray Spectrometer, Neutron Spectrometer, and High Energy Neutron Detector Experiment Data Records	ODY	View Information	Product Search	<ul style="list-style-type: none"> • 2001 Mars Odyssey Data Archives
2. 2001 Mars Odyssey Thermal Emission Imaging System Infrared Experiment Data Records	ODY	View Information	Product Search	<ul style="list-style-type: none"> • Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
3. 2001 Mars Odyssey Thermal Emission Imaging System Visible Experiment Data Records	ODY	View Information	Product Search	<ul style="list-style-type: none"> • Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
4. 2001 Mars Odyssey Thermal Emission Imaging System Infrared Reduced Data Records	ODY	View Information	Product Search	<ul style="list-style-type: none"> • Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas
5. 2001 Mars Odyssey Thermal Emission Imaging System Visible Reduced Data Records	ODY	View Information	Product Search	<ul style="list-style-type: none"> • Correlative Search • 2001 Mars Odyssey Data Archives • Planetary Image Atlas

Custom Data Set Browser – THEMIS/GRS



Planetary Image Atlas



[NEW SEARCH](#) [ABOUT](#) [HELP](#) [FEEDBACK](#) [HOME](#)

[QUICK SEARCH](#)

[GEOMETRY](#)

[INSTRUMENT](#)

[FEATURES](#)

[TIME](#)

[Reset this Page](#) [Reset All](#)

MISSION NAME

☒ 2001_MARS_ODYSSEY

INSTRUMENT

HEND_SPECTRA

THEMIS_IR

THEMIS_VIS

	Min	Max	Valid Range
<u>CENTER LATITUDE</u>	<input type="text"/>	<input type="text"/>	-90.0 – 90.0
<u>CENTER LONGITUDE</u>	<input type="text"/>	<input type="text"/>	0.0 – 360.0

[Preview](#) [Search ▶](#)

Web Page Curator: [Pamela Woncik](#), PDS Imaging Node, JPL

Default Data Set Browser – Radio Science



Basic Product Search

- ☒ Mars Global Surveyor Accelerometer EDR Data Set
- ☐ Mars Global Surveyor Accelerometer Altitude Data Set
- ☐ Mars Global Surveyor Accelerometer Profile Data Set

SOFTWARE, DOCUMENTATION AND
OTHER ANCILLARY DATA



Search for individual data products by entering one or more of the parameters below, or for further information on the dataset, read the [Data Set Catalog](#) file.



Product ID:

Identifier assigned to a data product

Release ID:

Identifier for a data set release

All

Latitude:

Latitude on the planet's surface

Longitude:

Longitude on the planet's surface

Orbit Number:

Orbital revolution of the spacecraft

Solar Longitude:

Measure of season on a target body

	Minimum	Maximum	Valid Range
			-87 degrees - 61.22 degrees
			-171 degrees - 354 degrees
			1 - 1285
			-171 degrees - 354 degrees

Default Data Set Browser – Radio Science



This search found 3170 products
Products 1 - 10 are displayed

PREVIOUS ◀ 10 ▶ NEXT

Product	Information	Download	Select for Download <input type="radio"/> All <input type="radio"/> None
1. MGS_AERODYNAMIC_PROPERTIES Product description goes here.	View Label	Get Product	<input type="checkbox"/>
2. PERIAPSIS_ORBITAL_ELEMENTS_P0040 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
3. PERIAPSIS_ORBITAL_ELEMENTS_P0041 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
4. PERIAPSIS_ORBITAL_ELEMENTS_P0042 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
5. PERIAPSIS_ORBITAL_ELEMENTS_P0043 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
6. PERIAPSIS_ORBITAL_ELEMENTS_P0044 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
7. PERIAPSIS_ORBITAL_ELEMENTS_P0045 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
8. PERIAPSIS_ORBITAL_ELEMENTS_P0046 Product description goes here.	View Label	Get Product	<input type="checkbox"/>
9. PERIAPSIS_ORBITAL_ELEMENTS_P0047 Product description goes here.	View Label	Get Product	<input type="checkbox"/>

Miscellaneous

Spin-off Benefits from D01



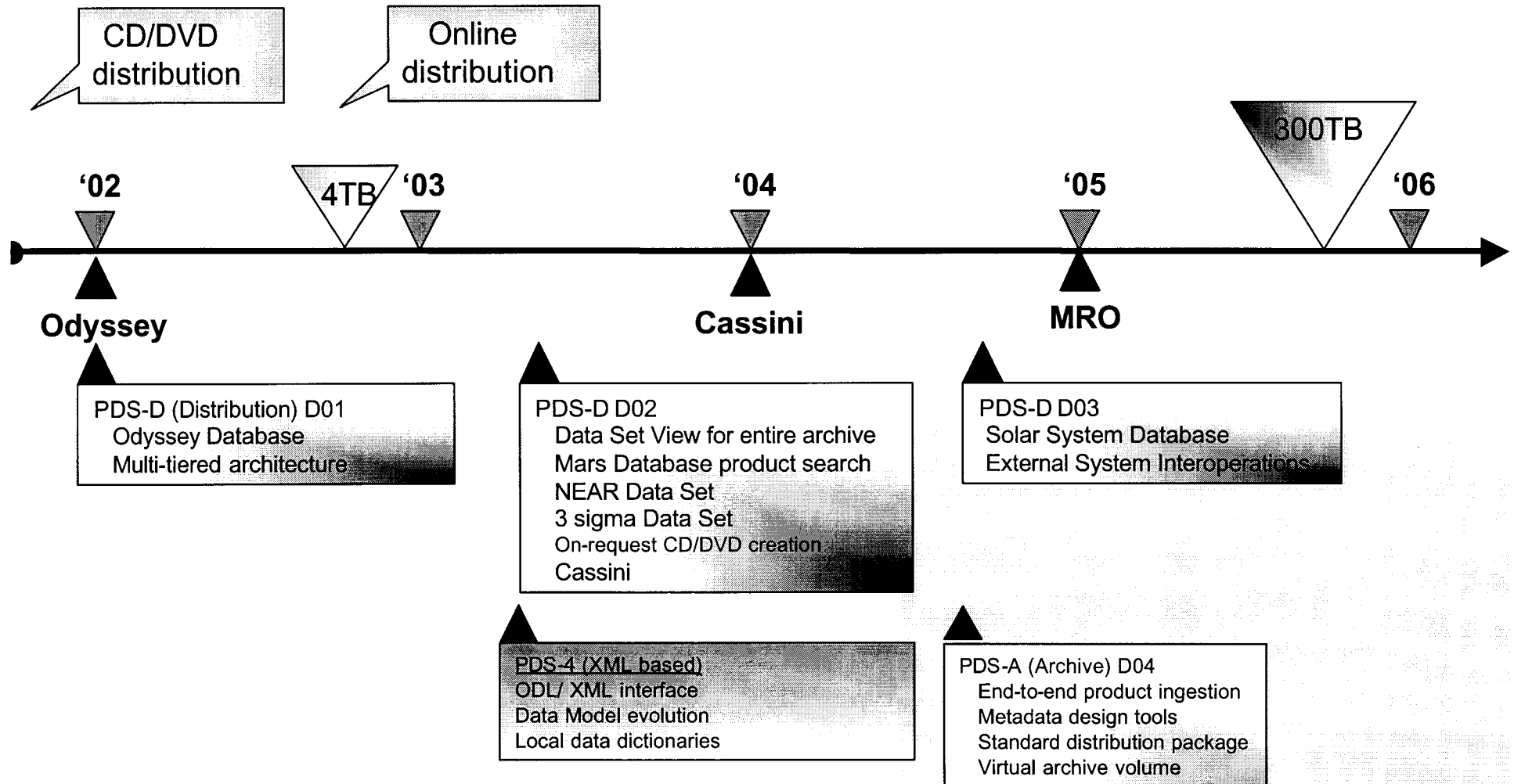
- **D01 infrastructure makes it easier to locate all online PDS resources**
- **New repository standard enables easier building of catalogs and cross-checking between catalogs and on-line archives**
- **Better tracking of data sets**
 - RELEASE_OBJECT will track data sets as data are accumulated
 - New paradigm of entering data sets into catalog as soon as possible
- **D01 required a data set catalog cleanup that benefited CN operations as well as development**
- **Catalog templates ingestion allowed as soon as possible**
 - No catalog templates exist on anybody's desk
 - Archive_status not = archived (all fields can be changed, including data_set_id)

Spin-off Benefits from D01 (cont)



- **Major system modifications caused minimal impact**
 - S wapped out CORBA for RMI
 - Migrated THEMIS data repository

PDS-D Development Timeline



Finally



- **Rapid development prototype proved concept**
 - Information architecture
 - Data architecture (data model)
 - Technology architecture (multi-tiered)
 - Used existing PDS resources and data model
 - Went from “ten years behind” to leading edge in 5 months
- **Multi-tiered architecture enables PDS-D and the PDS Vision**
 - *Seamless search and retrieval* of data products
 - Online access as primary method of *data distribution*
 - *Standard interfaces* for software developers
 - Provide real time access to *distributed heterogeneous* data repositories
 - Maintain *geographically distributed* data archives
 - Leverages current capabilities with minimal impact to existing resources
 - Plug-ins for *favorite tools*
- **PDS can evolve as basic information technology changes**

Acknowledgements





- Team Lead – Sue Lavoie
- OODT Team
 - Dan Crichton
 - Sean Kelly
 - Thuy Tran
 - Shu Lui
 - Jerry Crichton
- PDS CN
 - Ron Joyner
 - Joel Wilf
 - Cynthia Hall-Atkinson
 - Steve Hughes
- PDS Discipline Nodes
 - Pam Woncik
 - Elizabeth Duxbury-Rye
 - Susie Slavney
 - Tom Stein
 - Noel Gorelick
 - Lyle Hubner
 - Todd King
 - Patty Garcia
 - Lee Elson
 - Steve Levoe

Backup

Requirements on Middleware (cont)





Search

Home · Documents · Downloads · Your Account

2002-10-26

Menu

- [Home](#)
- [AvantGo](#)
- [Downloads](#)
- [FAQ](#)
- [Feedback](#)
- [Glossaries](#)
- [Journal](#)
- [Members List](#)
- [Private Messages](#)
- [Recommend Us](#)
- [Search](#)
- [Statistics](#)
- [Submit News](#)
- [Topics](#)
- [Web Links](#)
- [Your Account](#)

Glossaries

- [Metadata](#)
- [OODT](#)
- [Standards](#)
- [Middleware](#)

Language

Select Interface Language:

English

Object Oriented Data Technology is ...


- ... **Distributed middleware** architecture
- ... Resource discovery through **metadata**
- ... **Information** exchange through **standards**

[More about OODT](#)

Testing OODT in a Clustering Environment

Posted by [kelly](#) on Monday, October 28 @ 04:51:05 PST (26 reads) ([Read More...](#) | 14284 bytes more | [comments?](#) | 3 |)
Score: 0)


Alex Leung set up a Linux cluster to run the OODT software for the PDS single point of entry. Because the OODT software uses RMI which records object references based on IP addresses, we needed to make certain that the OODT software would function in such an environment. This article describes the results of such testing.



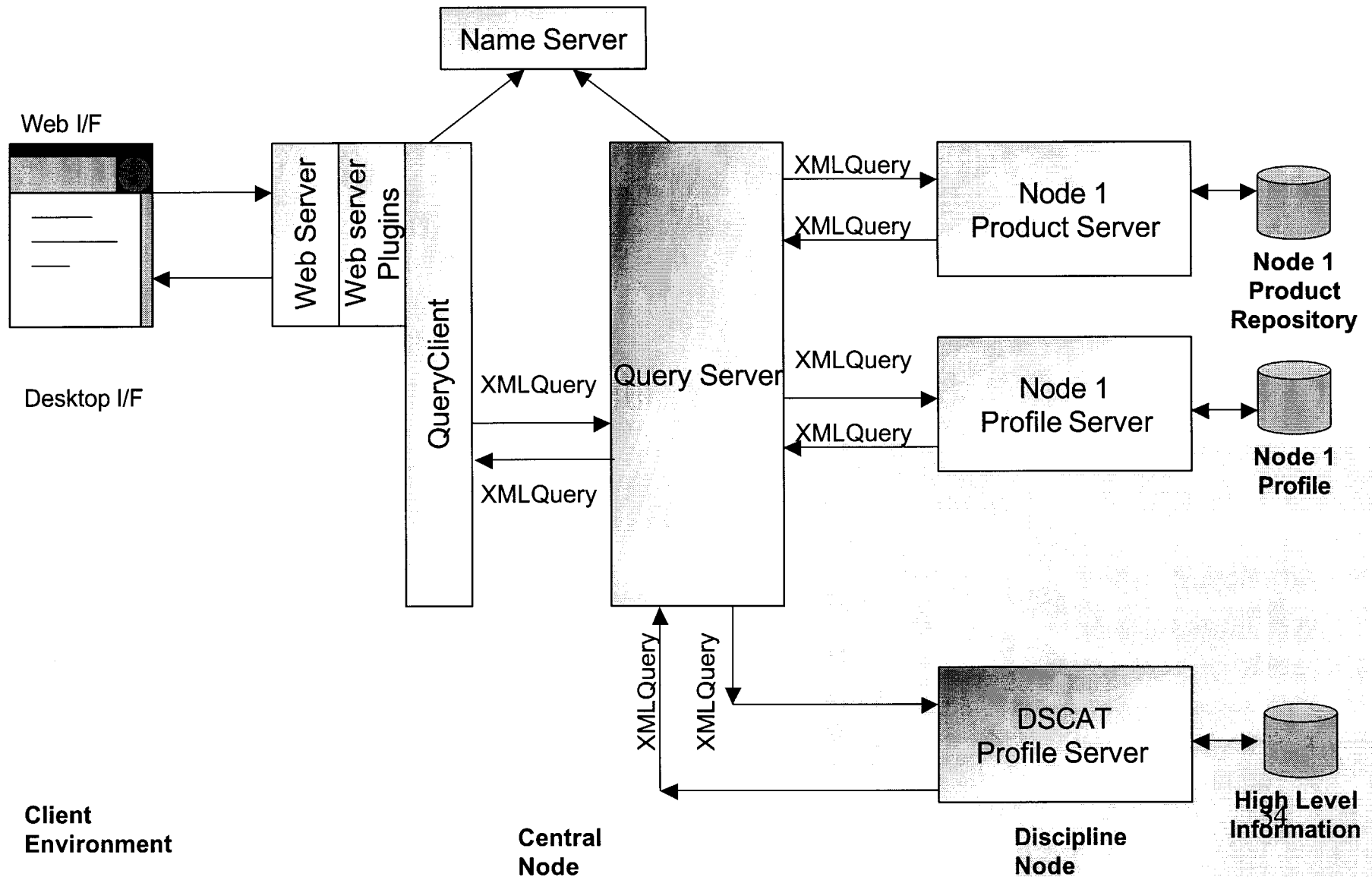
Using the Control Program

Posted by [kelly](#) on Tuesday, October 22 @ 14:19:34 PDT (6 reads) ([Read More...](#) | 18680 bytes more | [comments?](#) | 3 |)
Score: 0)

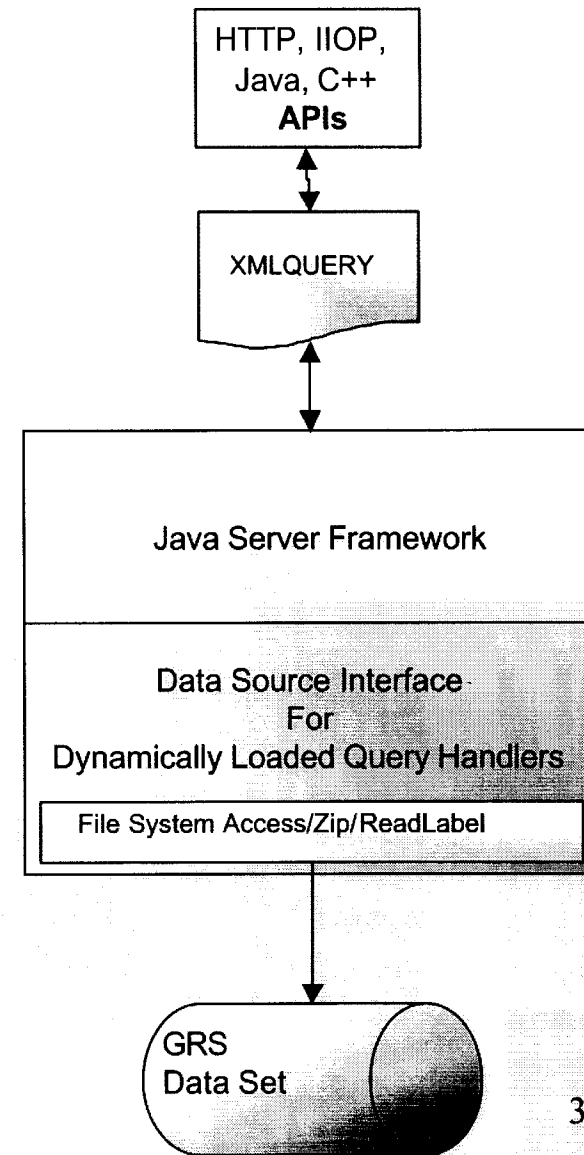
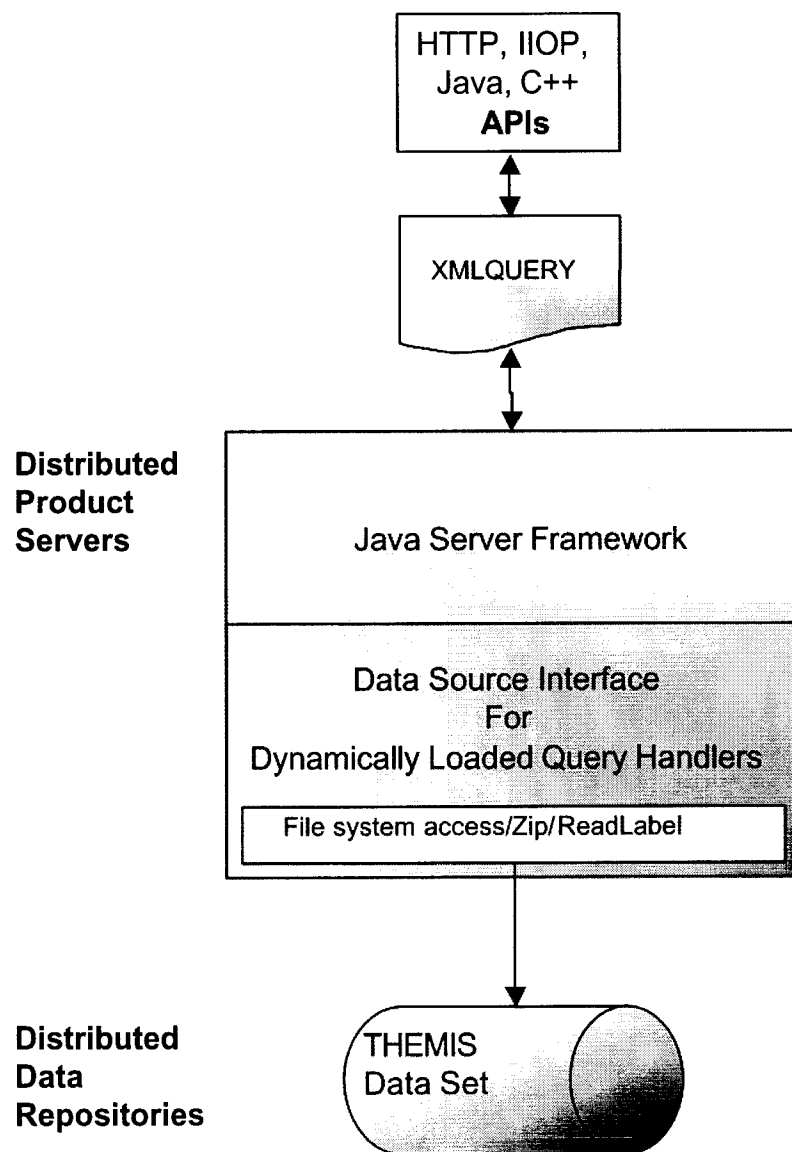
The command-line Control program provides a convenient way to control the entire network of OODT servers. This program, which is another Server Manager client, enables you to start and stop the entire network, install new files, and get status.



Conceptual PDS Implementation



Product Server Architecture



Scalability



- **Number of system component *interconnections increases linearly***
 - N nodes added as needed
 - One-to-one connections from each component to middleware
 - Exponential number of inter-operational connections made dynamically via message passing
- **Since distribution system is built as a *light layer* on top of the archive system, it will scale as long as the archive system scales**
- **Continue to distribute archive as needed to support larger data repositories (e.g. MRO)**
 - Parallel load balancing
- **Smaller frequently used data repositories can be mirrored**

Correlative Search the Simple Way



- **All data resources in the system are profiled**
- **Submit a query that describes what you want**
 - *Not how* to get what you want
- **System returns all matching data profiles**
 - Provides identification and description information
 - Provides location information
 - Provides all PDS metadata to support correlative science
 - Information is machine and human readable
- **Submit query to retrieve data**

PDS-D D01 Architecture



- **The architecture provides *seamless access to distributed data repositories and catalogs***
 - Location independence (e.g. the data appears local)
 - Information hiding (e.g. standard interfaces to differing catalogs and repositories)
 - Data driven architecture (i.e. use PDS data architecture)
 - Scalable and extensible (e.g. Odyssey through MRO)
 - Client APIs for search and retrieval of data and metadata
- **Maintains *geographically distributed data archives***
 - Conform to CODMAC recommendation for discipline data systems
 - Keeps data in the hands of the scientific experts
 - Promote closer ties with mission instrument teams
- **The architecture has *minimum impact on existing PDS resources***
 - Maintain original bits and convert as necessary
 - Leverage development using existing system resources
- **Separates data architecture from the technology architecture**